



Semnan University

Journal of Modeling in Engineering

Journal homepage: <https://modelling.semnan.ac.ir/>

ISSN: 2783-2538



Research Article

Applying Dictionary Learning Algorithms In Sparse Representation of Speech Signals

Naser Sharafi ^a, Salman Karimi ^{b,*}, Samira Mavaddati ^c

^a M.Sc. Student, Faculty of Engineering, Lorestan University, Khorramabad, Iran.

^b Assistant Professor, Faculty of Engineering, Lorestan University, Khorramabad, Iran.

^c Assistant Professor, Faculty of Engineering and Technology, University of Mazandaran, Babolsar, Iran

PAPER INFO

Paper history:

Received:

Revised:

Accepted:

Keywords:

Sparse representation;

Dictionary learning;

Speech processing;

K-SVD;

OMP;

STFT.

ABSTRACT

As a widely used technique in signal processing, Sparse representation has gained significant attention in various fields, including data compression, noise reduction in speech and image signals, pattern recognition, and other signal processing-related issues. In such representations, signals are linearly combined using a small number of dictionary atoms, leading to data dimensionality reduction and improved signal processing efficiency. To accurately represent speech data, an appropriate dictionary is required to effectively represent speech signals' characteristics. In this paper, dictionaries are trained using dictionary learning algorithms and sparse representations such as MOD, K-SVD, RAMC, UD4-MOD, and OMP, in the time, time-frequency, and wavelet transform domains. The performance of the obtained dictionaries is evaluated using various time-frequency metrics such as RE, MSE, fwSegSNR, SegSNR, PESQ, and STOI. The results demonstrate that employing the K-SVD dictionary learning algorithm in conjunction with the OMP sparse representation algorithm in the STFT domain achieves promising results for speech signal reconstruction.

DOI: <https://doi.org/>

© 2024 Published by Semnan University Press.

This is an open access article under the CC-BY 4.0 license.(<https://creativecommons.org/licenses/by/4.0/>)

* Corresponding author.

E-mail address: karimi.salman@lu.ac.ir

How to cite this article:

بکارگیری الگوریتم‌های یادگیری واژه‌نامه در بازنمایی تُنک دادگان گفتاری

ناصر شرفی^۱، سلمان کریمی^{۲*} و سمیرا مودتی^۳

اطلاعات مقاله	چکیده
دریافت مقاله:	بازنمایی تُنک به عنوان یکی از روش‌های پر کاربرد در پردازش سیگنال، در زمینه‌های مختلفی مانند فشرده‌سازی داده، حذف نویز از سیگنال‌های گفتاری و تصویری، تشخیص الگو و سایر مسائل مرتبط با پردازش سیگنال مورد توجه قرار گرفته است. در چنین بازنمایی‌هایی، سیگنال‌ها با استفاده از تعداد کمی از اتم‌های واژه‌نامه به صورت خطی ترکیب می‌شوند که منجر به کاهش ابعاد داده و بهبود کارایی در پردازش سیگنال می‌شود. به منظور بازنمایی دقیق‌تر داده‌های گفتاری، نیاز به واژه‌نامه مناسبی است که بتواند ویژگی‌های سیگنال گفتار را به خوبی نمایش دهد. در این مقاله، واژه‌نامه‌هایی با استفاده از الگوریتم‌های یادگیری واژه‌نامه و بازنمایی تُنک K-MOD، K-SVD، RAMC، SVD و UD4-MOD و بازنمایی تُنک OMP در حوزه‌های زمان، نمایش زمان-فرکانس و تبدیل موجک آموزش داده می‌شوند. ارزیابی کارایی واژه‌نامه‌های به دست آمده با استفاده از معیارهای مختلف زمانی و فرکانسی مانند RE، MSE، fwSegSNR، SegSNR، PESQ و STOI انجام شده است. نتایج حاصل، نشان می‌دهد که بکارگیری الگوریتم یادگیری واژه‌نامه K-SVD در ترکیب با الگوریتم بازنمایی تُنک OMP در حوزه STFT نتایج مطلوبی را به منظور بازسازی سیگنال گفتاری به دست می‌دهد.
بازنگری مقاله:	
پذیرش مقاله:	
واژگان کلیدی:	
بازنمایی تُنک،	
آموزش واژه‌نامه،	
پردازش سیگنال،	
K-SVD	
OMP	
تبدیل فوری کوتاه مدت.	

DOI: <https://doi.org/>

© 2024 Published by Semnan University Press.

This is an open access article under the CC-BY 4.0 license. (<https://creativecommons.org/licenses/by/4.0/>)۱- مقدمه^۱

به تصویر می‌کشد و می‌توان از آن برای نمایش داده‌های جدید و نادیده به صورت تُنک و کارآمد استفاده نمود. بازنمایی تُنک به روال نمایش یک بردار به عنوان ترکیبی از چند اتم واژه‌نامه اشاره دارد. با انتخاب تنها چند بردار از واژه‌نامه آموخته شده، می‌توان تعداد زیادی نمونه را در فضایی با ابعاد کم نمایش داد. این مفاهیم برای اولین بار در سال ۲۰۰۶ با عنوان حسگری فشرده^۴ معرفی گردید که کمک شایانی در توسعه نظریه بازنمایی تُنک داشته

یادگیری واژه‌نامه^۲ و بازنمایی تُنک^۳، دو روش مرتبط در یادگیری ماشین هستند که اغلب برای پردازش و تجزیه و تحلیل مجموعه داده‌های بزرگ کارایی دارند. یادگیری واژه‌نامه شامل یافتن مجموعه‌ای از توابع اتم یا پایه به نام واژه‌نامه است که می‌تواند تعداد زیادی نمونه آموزشی را بازنمایی نماید. واژه‌نامه آموخته شده مجموعه‌ای از بردارهایی است که ویژگی‌های کلیدی داده‌های آموزشی را

* پست الکترونیک نویسنده مسئول: karimi.salman@lu.ac.ir

۱. کارشناسی ارشد، دانشکده فنی و مهندسی، دانشگاه لرستان

۲. استادیار، دانشکده فنی و مهندسی، گروه مهندسی برق الکترونیک، دانشگاه

لرستان

۳. استادیار، دانشکده مهندسی و فناوری، دانشگاه مازندران

استاد به این مقاله: نحوه استناد فارسی در اینجا درج گردد.

2 Dictionary Learning

3 Sparse Representation

4 Compressive Sensing

است [۱]. کاربرد نمایش تُنک به تدریج از فشرده‌سازی داده‌های اولیه به بسیاری از زمینه‌های پردازش تصویر و گفتار، پردازش زبان طبیعی و سیستم‌های توصیه‌گر گسترش یافت. برخی از این زمینه‌ها شامل حذف نویز، تشخیص الگو، تفسیر تصاویر پزشکی و پردازش گفتار می‌باشد که در آن‌ها داده‌ها به صورت ترکیبی از تعداد کمی از اجزاء پایه مدل می‌شوند [۵-۲]. در این مقاله کاربردهای این دو مفهوم به منظور بازنمایی سیگنال گفتار مورد بررسی قرار می‌گیرد. در این رابطه تاکنون، مطالعات متنوعی انجام شده است که در ادامه به بررسی آن‌ها پرداخته می‌شود. با استفاده از بازنمایی تُنک و الگوریتم‌های یادگیری واژه‌نامه در [۶]، با هدف شناسایی نواحی غیر طبیعی مغز مبتلا به بیماری اسکیزوفرنی، واژه‌نامه جدیدی با ویژگی‌های مهم فشرده‌گی گروهی^۱ و غیرهمدوسی^۲ میان اتم‌های آن طراحی شده است. نتایج بر روی تصاویر مغزی نشان‌دهنده دقت و حساسیت بالای این روش در شناسایی نواحی مهم عصبی می‌باشد. در [۷]، تصاویر با وضوح طیفی بالا (مانند تصاویر فراطیفی) با تصاویر با وضوح فضایی بالا ترکیب می‌شوند. این پژوهش براساس استراتژی ترکیب بیزین بوده که از اطلاعات پیشین و یادگیری واژه‌نامه برای بازنمایی نسبت داده به تصویر هدف که وضوح طیفی و فضایی بالایی دارد، استفاده می‌نماید. همچنین در این روش به توسعه روش‌های جدید تخمین حرکت بیزین برای تصاویر سونوگرافی پرداخته شده است. در [۸]، یک الگوریتم مبتنی بر یادگیری واژه‌نامه برای بهبود کیفیت گفتار از طریق بازنمایی تُنک در حوزه تبدیل بسته موجک معرفی شده است. در این روش، یادگیری واژه‌نامه تُنک برای داده‌های آموزشی گفتار و نویز براساس معیار همدوسی در هر زیرباند پیشنهاد شده است. این الگوریتم‌ها همدوسی درونی واژه‌نامه‌ها و همدوسی متقابل واژه‌نامه‌های گفتار و نویز را کاهش می‌دهد. الگوریتم در هر دو سناریوی نظارت‌شده و نیمه‌نظارت‌شده ارائه گردیده است [۸]. در [۹]، الگوریتم جدیدی به نام بازنمایی تُنک در دامنه طیفی-زمانی^۳ (STRF) برای تشخیص فعالیت صوتی^۴ (VAD) ارائه شده است. این الگوریتم براساس مدل قشر شنوایی برای استخراج ویژگی و مدولاسیون طیفی-زمانی چند مقیاسی برای طبقه‌بندی

صدا طراحی شده است. این الگوریتم گفتار را به حوزه طیفی-زمانی^۵ انتقال داده و از ویژگی‌های مبتنی بر شنوایی را با مقیاس‌های متعدد تفکیک زمانی و طیفی به دست می‌دهد. یک روش نوآورانه به منظور بهبود کیفیت گفتار با استفاده از نمایش تُنک و مقیاس فرکانسی لگاریتمی در [۱۰] ارائه شده است. این روش با کاهش اندازه واژه‌نامه و بدون نیاز به تخمین نویز می‌تواند به طور مؤثری در شرایط نویز متغیر با نسبت‌های پایین سیگنال به نویز موثر بوده و کیفیت سیگنال گفتار را بهبود بخشد. با توجه به نتایج به دست آمده، این روش قابلیت اجرا در شرایط سخت محیطی را به اثبات رسانده و به عنوان یک پیشرفت قابل توجه در عرصه بهبود کیفیت گفتار مطرح می‌شود. یک الگوریتم جدید یادگیری واژه‌نامه مختص بهبود کیفیت گفتار در محیط‌هایی با نویز غیرایستا در [۱۱] معرفی شده است. این روش با تمرکز بر کاهش همدوسی بین زیرواژه‌نامه‌های گفتار و نویز و ارزیابی دقیق خطاهای مرتبط با اعوجاج منابع^۶، در مرحله بهسازی از کدگذاری تُنک استفاده می‌نماید تا تخمینی از طیف دامنه گفتار و نویز فراهم نماید. استفاده از فیلتر وینر در مرحله نهایی به اصلاح و بهبود بیشتر کیفیت گفتار کمک می‌نماید. در [۱۲]، یک روش جدید بهبود کیفیت گفتار مبتنی بر یادگیری واژه‌نامه و تجزیه ماتریس رتبه پایین^۷ پیشنهاد شده است. این روش با ارائه مدلی جدید از نویز و ترکیب آن با ماتریس کم‌رتبه و واژه‌نامه فراکامل^۸، واژه‌نامه نویز را با همدوسی کمی نسبت به واژه‌نامه گفتار در مرحله یادگیری، آموزش می‌دهد. سپس مؤلفه کم‌رتبه نویز با الگوریتمی تحت‌نظارت، استخراج و خوشه‌بندی می‌شود. در نهایت در مرحله بهبود، گفتار تمیز از گفتار نویزی بازیابی می‌گردد. نتایج آزمایش‌ها حاکی از برتری این روش نسبت به روش‌های پایه در این حوزه پردازشی مانند الگوریتم‌های مبتنی بر یادگیری واژه‌نامه برای بهبود کیفیت گفتار است که در [۱۲] معرفی شده‌اند. سیستم VAD طراحی شده در هر سناریو براساس انرژی ضرایب تُنک بکار گرفته شده است. هدف در [۱۳]، توسعه یک روش یادگیری واژه‌نامه فشرده است که خطای بازنمایی تُنک را به حداقل می‌رساند و به چارچوب‌های سخت نرم

⁵ Spectro-Temporal Domain⁶ Source Distortion²¹ Low Rank Matrix²² Overcomplete Dictionary⁵ Group Sparsity⁶ Incoherence³ Sparse Representation in Spectro-Temporal Domain⁴ Voice Activity Detection

واحد^۱ (UNTF) نزدیک می‌شود. این امر با یک رویکرد مبتنی بر گرادیان برای به دست آوردن واژه‌نامه فشرده که در آن محدودیت‌های عادی‌سازی در تابع هزینه با استفاده از یک مقداردهی اولیه احتمالی طراحی گردیده، حاصل شده است. واژه‌نامه فشرده آموخته‌شده به منظور بازیابی گفتار آسیب دیده بکار گرفته می‌شود تا سیگنال‌های گفتاری دارای اعوجاج توسط داده‌های از دست رفته را بازیابی نماید. در [۱۴]، روش جدیدی براساس تمایز بین آوای صدادار و بی‌صدا در جملات پیشنهاد گردیده است. به منظور پیاده‌سازی این روش، نویسندگان الگوریتم تعقیب تطبیق متعامد^۲ (OMP) را برای دستیابی به نمایش پراکنده انتخاب نموده و سپس الگوریتم تجزیه K مقدار تکین^۳ (K -SVD) برای ساخت واژه‌نامه در نظر گرفته شده است. همانطور که بیان شد الگوریتم K -SVD به دلیل توانایی در مدیریت مجموعه داده‌های بزرگ و با ابعاد بالا می‌تواند در این راستا مفید باشد. در این روش، تعداد کافی از آوای صدادار و بی‌صدا را برای یادگیری واژه‌نامه انتخاب و از K -SVD برای ساخت واژه‌نامه‌های مخصوص به هر دسته استفاده گردید. سپس، سیگنال‌های مورد ارزیابی به صورت نمایش تُنک در هر دو واژه‌نامه بیان می‌گردند. شیوه تفکیک آوای صدادار یا بی‌صدا در جمله، با مقایسه پراکندگی ضرایب در دو واژه‌نامه انجام و نتایج بر اثربخشی این روش تأیید می‌نماید. در مقاله [۱۵]، روش جدیدی با نام بازنمایی تُنک خوشه‌بندی‌شده^۴ (CSSR) معرفی شده که به بهبود استقلال صدای گوینده در تبدیل صدا^۵ کمک می‌نماید. این روش با ترکیب دو جزء یادگیری واژه‌نامه ساختار یافته خوشه‌ای و تابع هدف انتخابی خوشه‌ای، کارایی واژه‌نامه و نمایش تُنک را بهبود بخشیده و در نتیجه، عملکرد تبدیل صدا را افزایش می‌دهد. در [۱۶]، الگوریتم بهسازی گفتار مبتنی بر حسگری فشرده پیشنهاد شد که در آن نویز در مرحله اندازه‌گیری حذف شده و بازیابی به صورت فشرده انجام شده است. این روش از آموزش واژه‌نامه همراه با الگوریتم K -SVD برای ایجاد یک واژه‌نامه فراکامل استفاده می‌نماید که قادر به تخمین نویز در قاب‌های سکوت است. به منظور طبقه‌بندی قاب‌های گفتاری و سکوت، از

الگوریتم VAD و یک تابع ماسک براساس خروجی این الگوریتم استفاده می‌شود. سپس، بردار مشاهده شده نویز تخمینی از بردار گفتار فعال کم می‌شود و با استفاده از بازیابی فشرده، طیف گفتار بهبود یافته به دست می‌آید. روشی جدید برای بهسازی سیگنال گفتاری به کمک روش‌های بازنمایی تُنک و یادگیری واژه‌نامه در [۱۷] معرفی شده است. این روش از یک روال یادگیری ناهمدوس استفاده می‌نماید تا با استفاده از تعداد محدودی واژه‌نامه‌های یادگیری‌شده در فضای ویژگی‌های موجک، سیگنال نویز از سیگنال گفتاری حذف شود. با تجزیه سیگنال به زیرباندهای مختلف، دقیق‌ترین اطلاعات گفتاری به دست می‌آیند. در ادامه، روش‌های نظارت‌شده و نیمه‌نظارت‌شده برای بهسازی گفتار ارزیابی و الگوریتمی برای تشخیص VAD پایه‌گذاری شد تا در گام بعد منجر به بهسازی گفتار گردد. در [۱۸]، به منظور بهبود عملکرد سیستم‌های تشخیص گفتار^۶ (ASR)، سیگنال‌های گفتاری به بلوک‌های آکوستیکی مجزا تقسیم و با کمک تبدیل بسته موجک^۷ (WPT) به منظور پارامتری نمودن داده، گفتار برای تشخیص بهتر مرزهای واجی با استفاده از طبقه‌بند بازنمایی تُنک^۸ (SRC) نمایش داده می‌شود. در [۱۹]، از تبدیل فوریه گسسته^۹ (DFT) به منظور آموزش واژه‌نامه مختلط استفاده شده تا امکان بازنمایی تُنک با کیفیت بالا و پیچیدگی محاسباتی کمتری از سیگنال‌های صوتی فراهم شود. در [۲۰]، روش جدیدی برای الگوریتم OMP پیشنهاد شده تا کاربرد آن برای سیگنال‌های گفتاری با ساختار پیچیده بررسی شود. در این روش، واژه‌نامه مختلطی بر روی تبدیل فوریه گسسته مدل گردید و مکانیزم متعامدسازی جدیدی معرفی شد. این اصلاح، توانایی الگوریتم OMP را برای دستیابی به یک بازنمایی فشرده و با کیفیت بالا از سیگنال‌های گفتاری با پیچیدگی محاسباتی کم، افزایش داده است. همچنین یک روش طبقه‌بندی براساس نمونه‌ها و بازنمایی تُنک در [۲۱] ارائه شده تا از طریق اطلاعات گفتاری، بیماری پارکینسون شناسایی گردد. در این روش، استفاده از واژه‌نامه‌های ویژه هر کلاس به منظور افزایش کارایی روش‌های طبقه‌بندی بازنمایی تُنک که بر پایه

⁶ Automatic Speech Recognition

⁷ Wavelet Packet Transformation

⁸ Sparse Representation Classifier

⁹ Discrete Fourier Transform

¹ Unit-Norm Tight Frame

² Orthogonal Matching Pursuit

³ K-Singular Value Decomposition

⁴ Cluster-Structured Sparse Representation

⁵ Voice Conversion

حداقل سازی مربعات^۱ l_1 و حداقل سازی مربعات مثبت^۲ (NNLS) می باشند، پیشنهاد شد. هدف، یافتن نمایشی تُنک برای بردارهای ویژگی های گفتاری آزمایشی است که به بازنمایی نمونه های گفتاری داده های آموزشی ارتباط دارد. یکی از اصلی ترین مزایای استفاده از این روش، مقاومت بالای آن در برابر تکرار و نویز داده ها و عدم نیاز به تنظیمات پیچیده و زیاد پارامترهای فرابعدی است.

جدول ۱، گزیده ای از پژوهش های انجام شده در سال های اخیر براساس مفاهیم یادگیری واژه نامه و بازنمایی تُنک به منظور پردازش گفتار را نشان می دهد. بررسی های انجام شده در این جدول بر روی روش های ارائه شده براساس ویژگی های به کار رفته، اهداف تحقیقاتی، منابع داده ای و

جدول ۱: پژوهش های اخیر در زمینه پردازش گفتار براساس آموزش واژه نامه و بازنمایی تُنک.

مرجع	حوزه ویژگی	هدف	پایگاه داده مورد استفاده	سال انتشار
[۵]	ضرایب کپسترال فرکانس مل ^۳ (MFCC)	شناسایی گفتار	TIMIT, AURORA4	۲۰۱۷
[۸]	WPT	بهسازی گفتار	GRID	۲۰۱۷
[۹]	دامنه طیفی-زمانی	VAD	TIMIT, NOISEX_92	۲۰۱۸
[۱۰]	لگاریتم فرکانسی	بهسازی گفتار	NOISEX_92	۲۰۱۸
[۱۱]	تبدیل فوریه کوتاه مدت ^۴ (STFT)	بهسازی گفتار	NOIZEUS	۲۰۱۸
[۱۲]	STFT	بهسازی گفتار	NOISEX_92, GRID	۲۰۱۸
[۱۳]	STFT	بهسازی گفتار	NOISEX_92, GRID	۲۰۱۸
[۱۴]	حوزه زمان	طبقه بندی حروف صدادار و بی صدا	----	۲۰۱۹
[۱۵]	زمان	بهسازی تبدیل صدا	CMU ARCTIC	۲۰۲۰
[۱۶]	STFT	بهسازی گفتار	NOIZEUS, NOISEX_92	۲۰۲۰
[۱۷]	تبدیل موجک گسسته ^۵ (DWT)	بهسازی گفتار	GRID	۲۰۲۰
[۱۸]	WPT	تشخیص خودکار گفتار	TIMIT	۲۰۲۱
[۱۹]	DFT	بازسازی گفتار	TIMIT	۲۰۲۱
[۲۰]	DFT	بازنمایی تُنک	NOIZEUS, NOISEX_92	۲۰۲۲
[۲۱]	MFCC	تشخیص بیماری	PC-GITA, MDVR-KCL	۲۰۲۳

^۴ Short-Time Fourier Transform

^۵ Discrete Wavelet Trannsform

^۱ 11-Regularized Least Squares

^۲ Non-Negative Least Squares

^۳ Mel Frequency Cepstral Coefficient

و غیره ساخته شوند. این واژه‌نامه‌ها به دلیل محدودیت‌های خود ممکن است قادر به ارائه نمایش تُنک برای طیف گسترده‌ای از سیگنال‌ها نباشند. در این صورت واژه‌نامه‌ها می‌توانند براساس داده‌ها و الگوریتم‌های یادگیری آن آموزش داده شوند تا به نحو اثربخشی با داده‌های مورد مطالعه تطبیق پیدا نمایند. گام آغازین در یک الگوریتم یادگیری واژه‌نامه استفاده از یک حوزه تبدیل و سپس انتقال سیگنال گفتاری به آن فضای ویژگی جدید به منظور تحلیل مناسب می‌باشد. این حوزه در روش پیشنهادی، فضای تبدیل موجک، زمان و نمایش زمان-فرکانس در نظر گرفته شده که به صورت زیر قابل نمایش است:

$$S = DX \quad (۱)$$

در این رابطه S سیگنال گفتار، D واژه‌نامه مورد انتظار و X بیانگر ضرایب در بازنمایی تُنک می‌باشد. ماتریس داده ورودی $Y \in \mathbb{R}^{P \times F}$ می‌تواند با ترکیب خطی تُنکی از اتم‌ها که $D \in \mathbb{R}^{P \times L}, L > P$ یک واژه‌نامه فراکامل است نمایش داده شود. واژه‌نامه شامل L اتم در ستون‌ها $\{d_l\}_{l=1}^L$ با نُرم واحد $\|d_{(:,l)}\|_2 = 1, \forall l = 1, \dots, L$ و بردار گذار K -تُنک $X \in \mathbb{R}^{L \times F}, L \gg K$ شامل ضرایب بازنمایی خواهد بود [۲۲]. در حالت کلی مسئله بازنمایی تُنک که شامل بخش‌های خطای بازسازی و قید تُنکی می‌باشد به صورت زیر قابل بیان است [۲۳]:

$$X^* = \underset{X}{\operatorname{argmin}} \|Y - DX\|_F^2, \|X\|_0 \leq K \quad (۲)$$

تعداد ضرایب غیرصفر یا کارینالیتی^۳ سیگنال در ماتریس ضرایب تنک X به صورت $\|X\|_0 = K$ خواهد بود. نُرم l_0 در رابطه (۲) منجر به یک مسئله بهینه‌سازی غیرمحدب و NP-Hard می‌شود که ممکن است باعث گیر افتادن در مینیمم‌های محلی شود. جایگزینی نُرم l_0 با نُرم l_1 به عنوان یک روش آرام‌سازی محدب، حل مسئله را تسهیل کرده و احتمال گرفتار شدن در مینیمم‌های محلی را کاهش می‌دهد، اگرچه این جایگزینی لزوماً مسئله را کاملاً محدب نمی‌کند [۱۸].

$$X^* = \underset{X}{\operatorname{argmin}} \|Y - DX\|_F^2, \|X\|_1 \leq \tau \quad (۳)$$

شکل ۱ نمایشی از نحوه بازنمایی تُنک سیگنال‌ها را براساس چینش اتم‌ها نمایش می‌دهد.

• مقایسه عملکرد الگوریتم‌های یادگیری واژه‌نامه در حوزه‌های مختلف: عملکرد الگوریتم‌های یادگیری واژه‌نامه در حوزه‌های زمان، نمایش زمان-فرکانس و تبدیل موجک مورد بررسی قرار می‌گیرد.

• ارائه رویکردی جدید برای آموزش واژه‌نامه‌ها با استفاده از الگوریتم‌های یادگیری واژه‌نامه: این رویکرد می‌تواند به بهبود دقت بازنمایی تُنک داده‌های گفتاری در طیف وسیعی از اهداف پردازشی کمک نماید.

• ارائه نتایج تجربی که نشان‌دهنده برتری الگوریتم-K-SVD در ترکیب با OMP در حوزه STFT است: نتایج نشان می‌دهد که این ترکیب، دقت بازسازی سیگنال گفتاری را به طور قابل توجهی افزایش می‌دهد.

• ارائه معیارهای ارزیابی مناسب برای سنجش کارایی الگوریتم‌های یادگیری واژه‌نامه: از معیارهای مختلف زمانی و فرکانسی برای ارزیابی عملکرد الگوریتم‌ها استفاده شده است.

• شناسایی زمینه‌های تحقیقاتی آینده در زمینه یادگیری واژه‌نامه برای بازنمایی تُنک داده‌های گفتاری: این مقاله زمینه‌های تحقیقاتی بالقوه‌ای را برای بهبود بیشتر این تکنیک‌ها، مانند استفاده از یادگیری عمیق و تکنیک‌های فشرده‌سازی تطبیقی، برجسته می‌نماید.

در بخش دوم مقاله، روش‌های رایج آموزش واژه‌نامه و بازنمایی تُنک مرور می‌شود. سپس در بخش سوم، الگوریتم پیشنهادی و معیارهای ارزیابی عملکرد آن ارائه می‌گردد. در بخش چهارم به بررسی نتایج بهسازی گفتار در حوزه‌ای که بهترین بازسازی گفتار را داشته، پرداخته می‌شود. در نهایت، در بخش پنجم به بحث در مورد نتایج و جمع‌بندی دستاوردهای تحقیق حاضر پرداخته می‌شود.

۲- یادگیری واژه‌نامه براساس بازنمایی تُنک

به منظور بهینه‌سازی روال بازنمایی تُنک و حصول خطای بازنمایی کمتر، ضروری است که واژه‌نامه متناسب با داده‌های مورد بررسی در دسترس باشد. واژه‌نامه‌هایی که مورد استفاده قرار می‌گیرند ممکن است به صورت مجموعه‌ای از توابع پیش آماده مانند تبدیل موجک، تبدیل فوریه، تبدیل گسسته کسینوسی^۱ (DCT)، تبدیل Gabor

³ Cardinality

¹ Discrete Cosine Transform

² K-Sparse

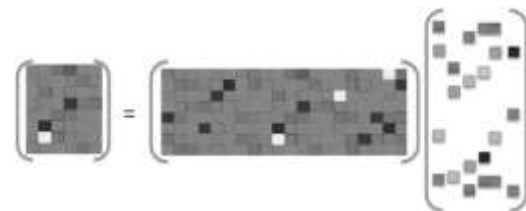
در سال‌های اخیر الگوریتم‌های زیادی برای تقریب ضرایب تُنک توسعه یافته‌اند. یکی از پرکاربردترین الگوریتم‌ها OMP می‌باشد. در این الگوریتم، هر مرحله با انتخاب اتمی انجام می‌شود که بیشترین همبستگی را با خطای باقی‌مانده که به صورت تفاوت سیگنال ورودی و ترکیب خطی اتم‌های قبلی محاسبه می‌گردد، داشته باشد. در مرحله اول از آنجایی که هیچ اتمی انتخاب نشده است، خطای باقی‌مانده معادل خود سیگنال ورودی در نظر گرفته می‌شود. علاوه بر این، معیارهای انتخاب بهینه برای بهترین اتم از خطای باقیمانده در تمام ستون‌های واژه‌نامه براساس حاصل ضرب درونی بردار خطای باقیمانده با هر ستون واژه‌نامه و به دنبال آن محاسبه ضریب همبستگی است. اتمی با بالاترین ضریب همبستگی به عنوان بهترین نماینده برای افزودن به اتم‌های قبلی، انتخاب و خطای باقیمانده بین سیگنال ورودی و تقریب فعلی را به حداقل می‌رساند. شرط توقف این الگوریتم تعیین حداکثر تعداد اتم‌ها در نمایش تُنک سیگنال یا رسیدن به حداقل خطای بازنمایی از پیش تعیین‌شده می‌باشد.

الگوریتم جهت‌های بهینه^۱ (MOD) یکی از روش‌های پایه در آموزش واژه‌نامه می‌باشد. در مرحله اول با استفاده از واژه‌نامه اولیه و یکی از الگوریتم‌های نمایش تُنک، ضرایب تُنک حاصل و از الگوریتم OMP در مرحله نخست یادگیری واژه‌نامه استفاده می‌شود [۲۵]. در مرحله بروزرسانی واژه‌نامه، هدف کاهش خطای بازنمایی کلی، $\|Y - DX\|_F^2$ است. برای این منظور، کافی است با استفاده از ماتریس ضرایب تُنک به‌دست‌آمده در مرحله‌ی قبل، مشتق خطای بازنمایی نسبت به D محاسبه و برابر صفر قرار داده شود. در نتیجه، یک رابطه برای واژه‌نامه‌ی به‌روزرشده‌ی D حاصل می‌شود:

$$D = YX^T(XX^T)^{-1} \quad (6)$$

پس از بروزرسانی واژه‌نامه، باید اتم‌های حاصل نرمالیزه شوند. همانطور که مشاهده می‌شود کل مجموعه اتم‌ها در این روش یک بار بروزرسانی می‌شوند که باعث سادگی پیاده‌سازی و سرعت اجرای الگوریتم می‌گردد. روش K-SVD یک از الگوریتم‌های کارآمد و مطرح در زمینه آموزش واژه‌نامه می‌باشد. تفاوت اصلی این روش با روال مبتنی بر MOD، در مرحله بروزرسانی واژه‌نامه است. در هر بار تکرار

در مسائل بهینه‌سازی بازنمایی تُنک، از نرم‌های مختلفی مانند نرم l_0 ، نرم l_1 ، نرم l_2 و نرم فروبینیوس استفاده می‌شود. نرم l_0 تعداد عناصر غیرصفر را شمارش می‌کند اما به دلیل غیرمحدب بودن، محاسبات پیچیده‌ای دارد. به همین دلیل، نرم l_1 به عنوان جایگزین محدب برای l_0 استفاده می‌شود تا از گیر افتادن در مینیمم‌های محلی جلوگیری شود. نرم l_2 بیشتر در مسائل مبتنی بر حداقل مربعات و برای کاهش حساسیت به نویز به کار می‌رود. همچنین، نرم فروبینیوس در مسائل مرتبط با ماتریس‌ها، مشابه نرم l_2 برای ماتریس‌ها، برای سنجش اختلاف بین داده‌ها و بازسازی‌ها استفاده می‌شود.



شکل ۱: بازنمایی تُنک سیگنال‌ها به صورت ترکیب خطی اتم‌های واژه‌نامه کدگذاری شده.

مسئله یادگیری واژه‌نامه را می‌توان به عنوان یک مسئله غیرمحدب توصیف نمود که به صورت زیر بیان می‌شود:

$$(D^*, X^*) = \underset{D, X}{\operatorname{argmin}} \|Y - DX\|_F^2 \quad (4)$$

برای حل مسئله فوق از روش کمینه‌سازی تناوبی^۱ استفاده می‌شود [۲۴]. روش کمینه‌سازی تناوبی شامل دو مرحله است. در گام اول ضرایب تُنک محاسبه می‌شوند و سپس در گام دوم بروزرسانی واژه‌نامه صورت می‌پذیرد. در مرحله اول با فرض ثابت در نظر گرفتن واژه‌نامه، نمایش تُنک برای دادگان آموزشی محاسبه و ضرایب تُنک بدست می‌آید. سپس در مرحله بعد با داشتن ضرایب تُنک، واژه‌نامه بروزرسانی شده و این روال تا رسیدن به خطای کمینه یا حداکثر تعداد تکرار، ادامه می‌یابد. این روال به صورت زیر تعریف می‌شود:

$$\begin{aligned} X^{(K+1)} &= \underset{X}{\operatorname{argmin}} \|Y - D^{(K)}X\|_F^2 \\ D^{(K+1)} &= \underset{D}{\operatorname{argmin}} \|Y - DX^{(K+1)}\|_F^2 \end{aligned} \quad (5)$$

تفاوت اکثر روش‌های یادگیری واژه‌نامه مربوط به تفاوت در این دو مرحله، به ویژه مرحله بروزرسانی واژه‌نامه می‌باشد.

¹ Alternating Minimization

می‌نماید [۲۷]. این روش نیز در هر مرحله بروزرسانی واژه‌نامه همانند MOD تمامی اتم‌ها را بروز می‌نماید. به دلیل ضرب داخلی D و X ، آموزش واژه‌نامه یک مسئله غیرمحدب است. در [۲۸]، این روش به صورت یک مسئله محدب ارائه شده که با جایگزینی عبارت محدب $X = X_0 + (X - X_0)$ و $D = D_0 + (D - D_0)$ در رابطه ۴ آموزش واژه‌نامه جدیدی طبق رابطه ۱۰ بیان می‌شود:

$$(D^*, X^*) = \underset{D, X}{\operatorname{argmin}} \|Y + D_0 X_0 - (D_0 X + D X_0)\|_F^2 \quad (10)$$

با ثابت نگاه‌داشتن پارامترهای مختلف در هنگام کمینه‌سازی تناوبی D و X ، چهار الگوریتم حاصل می‌شود. یکی از این الگوریتم‌ها که سرعت و دقت آن بر روی داده‌های مصنوعی نسبت به سایر الگوریتم‌های معرفی شده برتری دارد، کمینه‌سازی تناوبی به صورت رابطه زیر است:

$$\begin{aligned} R^{(k)} &= Y - (D^{(k)} - D^{(k-1)})X^{(k)} \\ X^{(k+1)} &= \underset{X}{\operatorname{argmin}} \|R^{(k)} - D^{(k-1)}X\|_F^2 \quad (11) \\ D^{(k+1)} &= \underset{D}{\operatorname{argmin}} \|Y - DX^{(k+1)}\|_F^2 \end{aligned}$$

گام بروزرسانی واژه‌نامه در این روش همان روال مورد استفاده در MOD می‌باشد. این روش با نام اختصاری UD4-MOD^۴ در [۲۸] معرفی شده است.

۳- بازنمایی سیگنال گفتار مبتنی بر یادگیری واژه‌نامه

سیگنال گفتار، اطلاعات پیچیده‌ای را در مورد گوینده، لحن و محتوای گفتار به همراه دارد. یکی از راه‌های پردازش و تحلیل این اطلاعات، روش‌های مبتنی بر بازنمایی سیگنال گفتار به صورت فشرده و دقیق است. به این منظور دسترسی به واژه‌نامه‌ای که اطلاعات دقیق و برجسته سیگنال گفتار در فضای ویژگی را مدنظر قرار دهد، ضروری است. در این بخش، به بررسی روش‌های مبتنی بر یادگیری واژه‌نامه برای بازنمایی سیگنال گفتار پرداخته می‌شود. این روش‌ها، واژه‌نامه را از خود داده‌های گفتاری می‌آموزد و از آن برای بازنمایی سیگنال گفتار به صورت ترکیبی خطی از اتم‌ها استفاده می‌کند.

این الگوریتم برخلاف MOD، اتم‌ها به صورت مجزا و همچنین سطر متناظر با این اتم در ماتریس ضرایب تُنک به طور متوالی بروز می‌شود. در این الگوریتم برای بروز کردن ستون k ام واژه‌نامه d_k ، سایر اتم‌ها ثابت در نظر گرفته می‌شوند و خطای باقیمانده مرتبط با هر اتم E_k ، از رابطه زیر محاسبه می‌شود [۲۶]:

$$\begin{aligned} \|Y - DX\|_F^2 &= \left\| Y - \sum_{i=1}^L d_i x^{[i]} \right\|_F^2 \\ \left\| Y - \sum_{i \neq k}^L d_i x^{[i]} - d_k x^{[k]} \right\|_F^2 &= \left\| E_k - d_k x^{[k]} \right\|_F^2 \quad (7) \end{aligned}$$

در این رابطه $x^{[i]}$ ، سطر i ام ماتریس X است. با اعمال تجزیه SVD بر روی E_k ، اتم d_k ضرایب تُنک مرتبط با اتم k ام $x^{[k]}$ بروز می‌شود. این اتم‌های بروز شده، برای بروزرسانی سایر اتم‌های باقیمانده نیز استفاده می‌شوند. الگوریتم RAMC^۱ یکی دیگر از روش‌های آموزش واژه‌نامه است که در سال‌های اخیر بر روی داده‌های آزمایشی تصادفی به خوبی پیاده‌سازی شده است [۲۷]. این روش کاهش همزمان همدوسی متقابل^۲ و میانگین همدوسی^۳ اتم‌های واژه‌نامه را نتیجه می‌دهد. این دو پارامتر به عنوان حداکثر اندازه همبستگی متقابل بین اتم‌ها $\mu(D)$ و میانگین اندازه همبستگی متقابل بین اتم‌ها $\mu_{avg}(D)$ تعریف شده و طبق رابطه زیر محاسبه می‌شوند [۲۷]:

$$\begin{aligned} \mu(D) &= \max_{i \neq j} |D^T D|_{ij} \\ \mu_{avg}(D) &= \sqrt{\frac{\|D^T D - I\|_F^2}{L(L-1)}} \quad (8) \end{aligned}$$

I ماتریس یک‌ه واحد و L تعداد اتم‌های واژه‌نامه را مشخص می‌نماید. گام بروزرسانی واژه‌نامه طبق رابطه زیر از طریق گرادیان نزولی محاسبه می‌شود [۲۷]:

$$D^{(k+1)} = D^{(k)} - \alpha \nabla_D F(D^{(k)}, X^{(k+1)}) \quad (9)$$

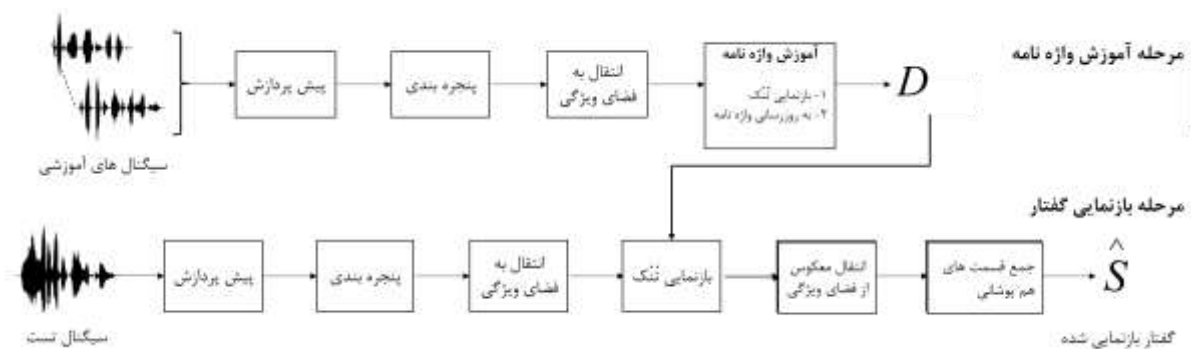
در این رابطه، F تابع هزینه و $\nabla_D F$ ، گرادیان آن نسبت به D را نشان می‌دهد. همچنین α ، شیب گرادیان نزولی را تنظیم

^۳ Average Coherence

^۴ UpDated4-MOD

^۱ Reduced Average and Mutual Coherence

^۲ Mutual Coherence



شکل ۲: بلوک دیاگرام مرحله آموزش واژه نامه و بازنمایی گفتار.

در شکل ۲، بلوک دیاگرام مرحله آموزش و بازنمایی سیگنال گفتار در فضای ویژگی مورد استفاده، نشان داده شده است. برای ارزیابی کارایی بازنمایی از معیارهای حداقل میانگین خطا^۱ (MSE)، خطای بازسازی^۲ (RE)، نسبت سیگنال به نویز بخش بندی شده^۳ (SegSNR)،^۴ (fwSegSNR)، ارزیابی ادراکی کیفیت گفتار^۵ (PESQ) و درک عینی کوتاه مدت^۶ (STOI) استفاده شده است. فضاهای ویژگی متداول مورد استفاده برای آموزش واژه نامه با توجه به کاربردهای مختلف سیگنال گفتار، حوزه های زمان، نمایش زمان-فرکانس و تبدیل بسته مویک و ... می باشد. واژه نامه گفتار در حوزه های ذکر شده براساس بازنمایی سیگنال های گفتار آموزش داده می شوند. به منظور بررسی روال بازنمایی، از گفتار پایگاه داده NOISEUES با فرکانس نمونه برداری ۸ کیلوهرتز در تمامی حوزه ها استفاده شده است. در ابتدا فضای ویژگی مطابق با روال بررسی شده در [۱۵] و [۱۴] در حوزه زمان در نظر گرفته می شود. با توجه به بلوک دیاگرام در شکل ۲، گفتار در مرحله آموزش با استفاده از پنجره های همپینگ با طول ۱۲/۵ میلی ثانیه و همپوشانی ۵۰٪ تقسیم بندی می شود تا داده های آموزشی برای آموزش واژه نامه در این حوزه ایجاد گردد. نرخ افزونگی واژه نامه های فراکامل به صورت تجربی و با توجه به کمترین خطای بازسازی در شبیه سازی ها، مقدار ۴ در نظر گرفته شده است [۸]. انتخاب این پارامتر به مقداری بیشتر از این مقدار، ضمن بالابردن زمان و هزینه محاسبات، تاثیر قابل توجهی

در بهبود نتایج شبیه سازی ندارد. همچنین، انتخاب عددی کمتر از این مقدار برای نرخ افزونگی، کاهش کیفیت و افزایش خطای بازنمایی را نتیجه می دهد. این پارامتر با در نظر گرفتن توازن بین زمان شبیه سازی و دقت نتایج بدست آمده، انتخاب شده است. شبیه سازی های انجام شده برای تعیین نرخ افزونگی براساس فضاهای ویژگی مختلف بکارگرفته شده و نیز روش آموزش واژه نامه K-SVD در جدول ۲ گزارش شده است. در جدول ۳، نتایج حاصل از آزمایش های مختلف در حوزه زمان با روش های مختلف آموزش MOD، K-SVD، RAM و UD4-MOD گزارش شده است. در [۸] و [۱۸]، WPT به عنوان فضای ویژگی برای آموزش واژه نامه بکار گرفته شده است. در این بخش، دو سناریو برای بازنمایی گفتار با WPT مورد بررسی قرار گرفته اند. در روش اول مطابق شکل ۲، سیگنال گفتار پس از پنجره گذاری، تا سطح ۳ با WPT تجزیه شده و هر قاب به ۸ زیرباند ضرایب بسته مویک تجزیه می شود. با داشتن این ۸ زیرباند ضرایب، ۸ واژه نامه مرتبط برای هر زیرباند آموزش داده می شود. در سناریو دوم، قبل از اعمال بلوک پنجره گذاری، WPT از گفتار کوتاه مدت حاصل و سپس، قاب بندی روی زیرباندهای ضرایب بسته مویک اعمال می گردد تا داده ها برای آموزش واژه نامه های مرتبط با هر زیرباند در دسترس باشند. در ادامه، ۸ واژه نامه برای استفاده در مرحله بازنمایی گفتار آموزش داده می شوند. جدول ۴ نتایج شبیه سازی ها برای دو سناریو فوق در فضای WPT را

⁴ Ferquency WSegmental SNR

⁵ Perceptual Evaluation of Speech Quality

⁶ Short Time Objective Intelligibility

¹ Mean Square Error

² Reconstruction Error

³ Segmental SNR

در مرحله‌ی آموزش، از دامنه‌ی حوزه STFT بهره‌برده شده است. این در حالی است که برای بازنمایی گفتار، فاز سیگنال بدون تغییر مورد استفاده قرار گرفته است. در جدول ۵، نتایج شبیه‌سازی‌های STFT گزارش شده که کارایی این روش در آموزش و بازنمایی گفتار را نشان می‌دهد.

نشان می‌دهد. استفاده از حوزه STFT به دلیل توانایی تحلیل و نمایش همزمان حوزه زمان و فرکانس سیگنال‌ها، یکی از انتخاب‌های پرکاربرد به منظور آموزش واژه‌نامه گفتار محسوب می‌شود. در پژوهش‌های [۱۱-۱۲]، [۱۳، ۱۶]، آموزش واژه‌نامه گفتار براساس تبدیل فوریه کوتاه‌مدت سیگنال گفتار انجام شده است. برای دستیابی به این هدف،

جدول ۲: تأثیر نرخ افزونگی بر خطای بازنمایی تُنک سیگنال گفتار براساس فضاهای ویژگی مختلف و هزینه محاسباتی.

نرخ افزونگی				معیارهای ارزیابی	فضای ویژگی
۸	۶	۴	۲		
۵۶	۴۲	۳۲	۲۱	زمان آموزش	زمان
۱۳/۹	۹/۳	۴/۶	۱/۰۶	زمان بازنمایی	
۰/۵۵	۰/۵۷	۰/۶۱	۰/۷۳	RE	
۱/۵e-۵	۱/۶e-۵	۱/۸e-۵	۲/۵e-۵	MSE	
۶۳	۴۹	۴۲	۲۶	زمان آموزش	WPT1
۷/۶	۲/۶	۲/۲	۱/۳	زمان بازنمایی	
۰/۵۱	۰/۵۲	۰/۵۳	۰/۵۷	RE	
۱/۳۰e-۵	۱/۳۳e-۵	۱/۳۷e-۵	۱/۵۸e-۵	MSE	
۷۰	۵۸	۴۳	۲۹	زمان آموزش	WPT2
۷/۸	۲/۴	۲/۲	۱/۳	زمان بازنمایی	
۰/۴۲	۰/۴۴	۰/۴۴	۰/۴۷	RE	
۰/۸۶e-۵	۰/۹۲e-۵	۰/۹۴e-۵	۱/۱e-۵	MSE	
۵۴	۴۲	۳۲	۲۱	زمان آموزش	STFT
۱۵/۹	۹/۳	۴/۶	۱/۱	زمان بازنمایی	
۰/۲۰	۰/۲۳	۰/۲۶	۰/۳۹	RE	
۰/۱۹e-۵	۰/۲۵e-۵	۰/۳۵e-۵	۰/۷۷e-۵	MSE	

جدول ۳: نتایج معیارهای ارزیابی بازنمایی گفتار در حوزه زمان.

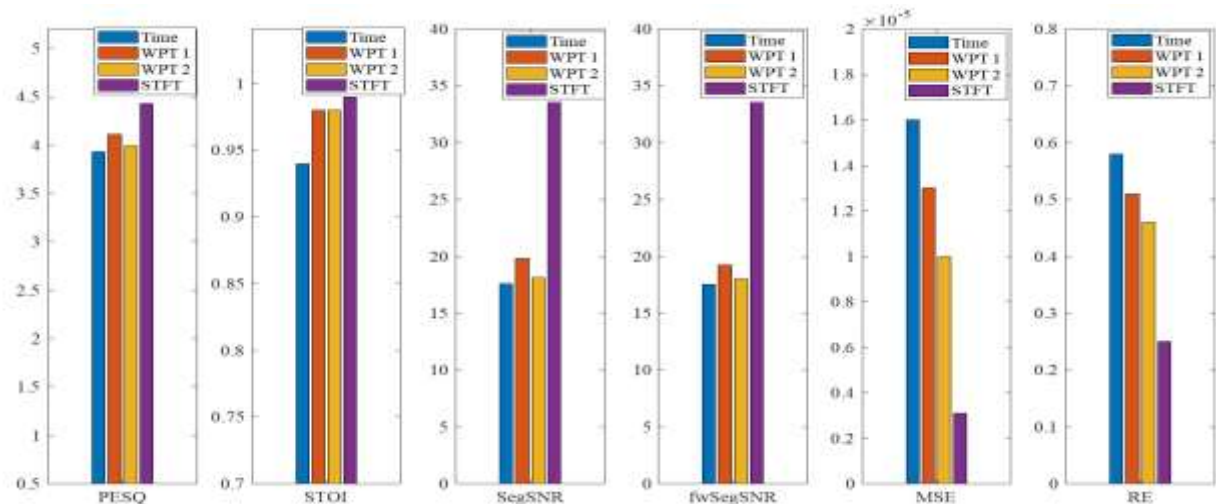
PESQ	STOI	fwSegSNR	SegSNR	RE	MSE	روش آموزش واژه‌نامه
۳/۸۷	۰/۹۳	۱۶/۸۴	۱۶/۸۳	۰/۶۲	۱/۸e-۵	MOD
۳/۸۷	۰/۹۴	۱۶/۸۹	۱۶/۸۸	۰/۶۰	۱/۷e-۵	K-SVD
۳/۲۸	۰/۹۲	۱۵/۳۹	۱۵/۵۹	۰/۸۸	۳/۸e-۵	RAMC
۳/۹۳	۰/۹۴	۱۷/۵۲	۱۷/۶۱	۰/۵۸	۱/۶e-۵	UD4-MOD

جدول ۴: نتایج معیارهای ارزیابی بازنمایی گفتار در حوزه WPT.

PESQ	STOI	fwSegSNR	SegSNR	RE	MSE	روش آموزش واژه‌نامه
۴/۱۱	۰/۹۸	۱۹/۲۳	۱۹/۸۳	۰/۵۱	۱/۳e-۵	روش ۱
۳/۹۹	۰/۹۸	۱۸/۰۱	۱۸/۱۳	۰/۴۶	۱e-۵	روش ۲
۴/۱۰	۰/۹۸	۱۸/۹۲	۱۹/۶۲	۰/۵۴	۱/۴e-۵	روش ۱
۳/۹۴	۰/۹۷	۱۷/۶۲	۱۷/۷۴	۰/۴۶	۱/۴e-۵	روش ۲
۳/۹۸	۰/۹۸	۱۸/۲۵	۱۹/۰۶	۰/۵۹	۱/۷e-۵	روش ۱
۳/۸۰	۰/۹۶	۱۶/۵۳	۱۶/۶۱	۰/۶۲	۱/۸e-۵	روش ۲
۳/۹۸	۰/۹۷	۱۷/۹۹	۱۸/۷۵	۰/۵۶	۱/۵e-۵	روش ۱
۳/۷۶	۰/۹۶	۱۶/۰۴	۱۶/۰۴	۰/۵۹	۱/۶e-۵	روش ۲

جدول ۵: نتایج معیارهای ارزیابی بازنمایی گفتار در حوزه STFT.

روش آموزش واژه‌نامه	MSE	RE	SegSNR	fwSegSNR	STOI	PESQ
MOD	۳/۲e-۶	۰/۲۵	۳۲/۹۸	۳۲/۹۲	۰/۹۸	۴/۱۴
K-SVD	۳/۱e-۶	۰/۲۵	۳۳/۵۶	۳۳/۴۷	۰/۹۹	۴/۴۳
RAMC	۳/۲e-۶	۰/۲۵	۳۳/۴۶	۳۳/۳۹	۰/۹۹	۴/۴۲
UD4-MOD	۷/۹e-۶	۰/۴۰	۳۰/۳۳	۳۰/۲۱	۰/۹۸	۴/۳۴



شکل ۳: مقایسه معیارهای ارزیابی بازنمایی گفتار براساس معیارهای مختلف ارزیابی گفتار در حوزه‌های مختلف بازنمایی.

همانطور که بیان شد بهسازی گفتار به عنوان یکی از کاربردهای بازنمایی و یادگیری واژه‌نامه در حوزه پردازش سیگنال شناخته می‌شود. در این بخش با توجه به کارایی بالا، تنها حوزه STFT نسبت به سایر حوزه‌ها در مسئله بازنمایی به عنوان فضای ویژگی در کاربرد بهسازی گفتار در نظر گرفته می‌شود. دامنه STFT سه نوع نویز سفید، ماشین و خیابان همانند گفتار تمیز توسط الگوریتم K-SVD آموزش داده شده است و با کنار هم قرار دادن واژه‌نامه گفتار بدون نویز D_s ، واژه‌نامه نویز D_n ، یک واژه‌نامه مرکب برای بررسی بهسازی متناسب با هر کدام از نویزها به صورت مجزا به صورت $D = [D_s, D_n]$ تشکیل می‌شود. سپس سیگنال‌های نویز با SNRهای ۵، ۰ و ۱۰ دسی‌بل به گفتار بدون نویز اضافه شده و مسئله بازنمایی با الگوریتم OMP با گفتار نویزی و واژه‌نامه مرکب بر طبق رابطه ۱۲ و ۱۳ دنبال می‌شود [۱۲]:

$$S_n \approx DX = [D_s, D_n] \begin{bmatrix} X_s \\ X_n \end{bmatrix}$$

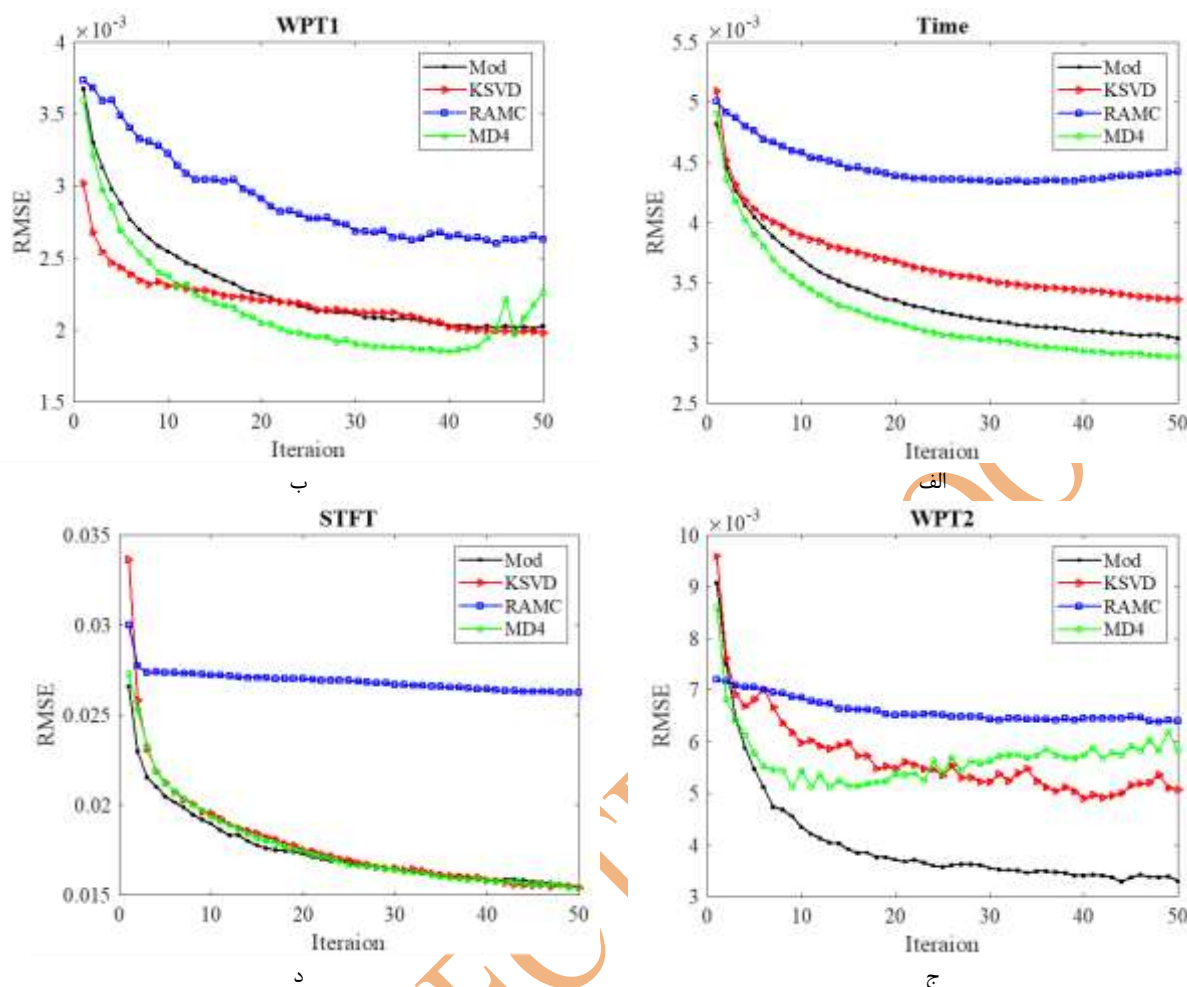
$$\hat{X} = \underset{X}{\operatorname{argmin}} \left\| S_n - [D_s, D_n] \begin{bmatrix} X_s \\ X_n \end{bmatrix} \right\|_F^2 \quad (12)$$

$$\hat{X} = \operatorname{OMP}(S_n, [D_s, D_n])$$

$$\hat{S}_s = D_s \hat{X}_s, \hat{N} = D_n \hat{X}_n \quad (13)$$

بررسی جداول ۳، ۴ و ۵ نشان می‌دهد که در حوزه زمان آموزش واژه‌نامه براساس روش UD4-MOD در مقایسه با سایر الگوریتم‌ها کارایی بیشتری دارد. در فضای ویژگی WPT به منظور بازنمایی گفتار، روش ۱ در مقایسه با روش ۲ عملکرد مطلوب‌تری دارد. بر طبق معیارهای ارزیابی، آموزش واژه‌نامه براساس الگوریتم MOD، در مقایسه با سایر روش‌های آموزشی مورد استفاده در این فضا موثرتر عمل می‌کند و همچنین الگوریتم K-SVD مورد استفاده برای آموزش واژه‌نامه در حوزه STFT نسبت به سایر روش‌ها کارایی بالاتری را به نمایش می‌گذارد. به منظور بررسی عمیق‌تر این مسئله که بازنمایی گفتار در کدام حوزه موفق‌تر بوده، نتایج برتر ذکر شده در هر فضای ویژگی برای همه معیارهای ارزیابی به عنوان شاخص آن حوزه در شکل ۳ نمایش داده شده است. همانطور که در نمودارهای این شکل مشاهده می‌شود، تمام معیارها گویای برتری قاطع حوزه STFT نسبت به سایر فضاهای مورد بررسی است. حوزه مربوط به روش اول فضای WPT و روش دوم WPT و در نهایت حوزه زمان در رتبه‌های بعدی قرار می‌گیرند.

۴- بهسازی گفتار براساس بازنمایی گفتار

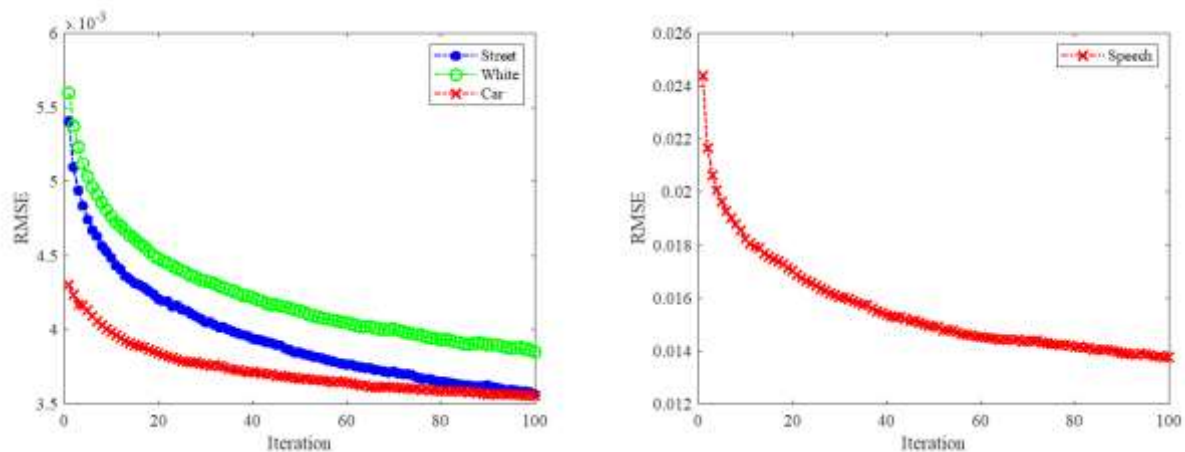


شکل ۴: منحنی‌های RMSE مرحله آموزش گفتار در حوزه (الف) زمان، (ب) WPT-روش اول، (ج) WPT-روش دوم، (د) STFT.

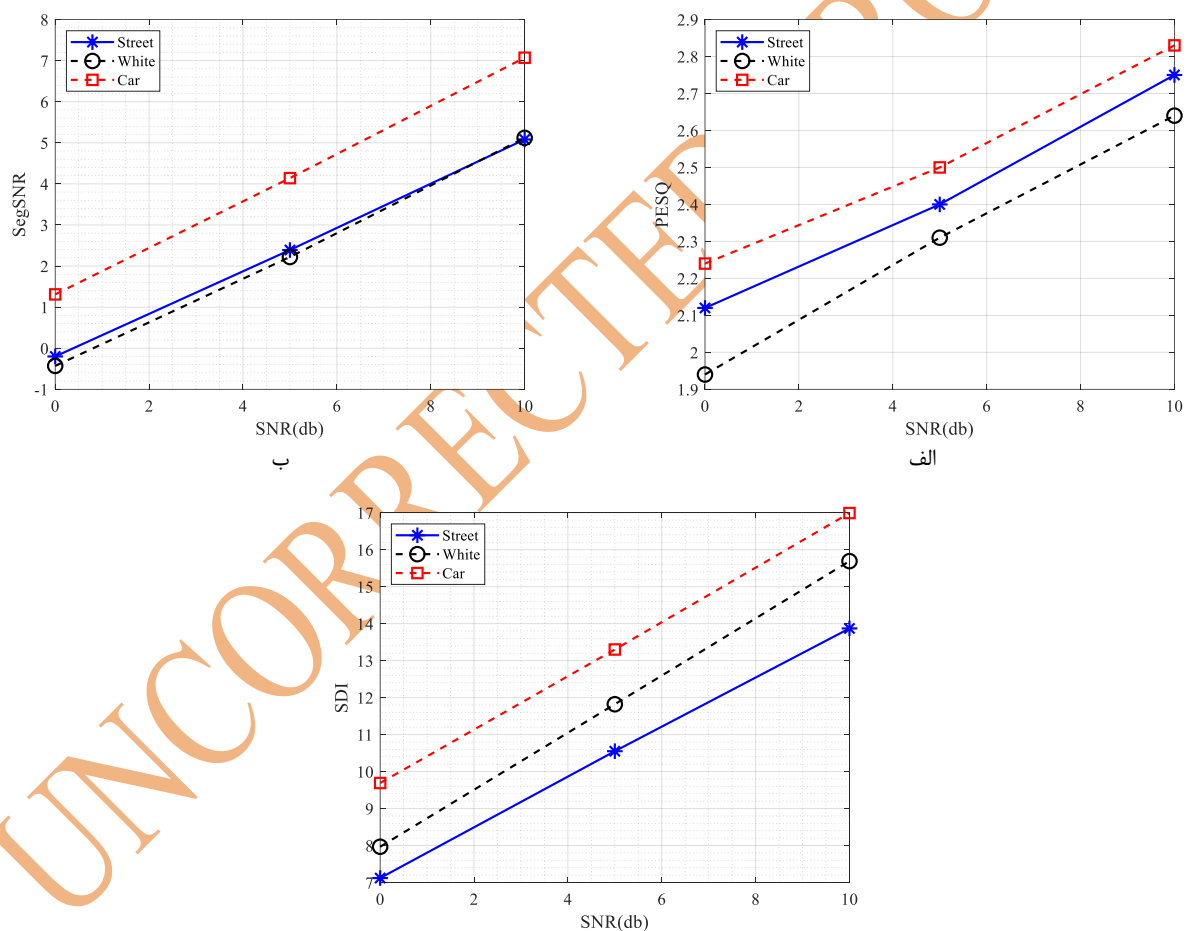
حاصل می‌شود. منحنی‌های RMSE مرحله آموزش گفتار در حوزه‌های مختلف، زمان، WPT-روش اول، WPT-روش دوم، و STFT در شکل ۴ نمایش داده شده است. همچنین شکل ۵، منحنی‌های RMSE بازسازی واژه‌نامه برای گفتار تمیز و نویز با الگوریتم آموزشی K-SVD را نشان می‌دهد. نرخ افزونگی در این آزمایش برای آموزش واژه‌نامه‌ها بر طبق نتایج گزارش شده در جدول ۲، برابر ۴ و مرحله تکرار الگوریتم ۱۰۰ در نظر گرفته شده است. در این مقاله سه معیار PESQ، SegSNR و SDI، برای ارزیابی کیفیت گفتار بهسازی شده با شرایط نویزی متفاوت در نظر گرفته شده‌است. نتایج بهسازی شده در نمودارهای شکل ۶ نشان داده شده‌است.

در روابط ۱۲ و ۱۳، S_n دامنه طیف گفتار نویزی، به صورت ترکیبی تُنک از واژه‌نامه گفتار تمیز و نویز در نظر گرفته شده است.

در مرحله بهسازی تلاش بر جداسازی اجزای گفتار تمیز و نویز با استفاده از واژه‌نامه مرکب است. X_n ، X_s زیرماتریس‌های X هستند که باید تخمین زده شوند. این کار با کمک الگوریتم بازنمایی OMP انجام می‌شود. $\hat{X}_s$ و $\hat{X}_n$ به ترتیب تخمینی از ماتریس ضرایب تُنک برای بازنمایی حدی اندازه طیف گفتار تمیز S_s و تخمین اندازه طیف نویز $\hat{N}$ ، نسبت به واژه‌نامه‌های مرتبط با آن‌ها یعنی D_n و D_s هستند. فیلتری شبیه فیلتر وینر بر روی دامنه‌های طیفی بدست آمده اعمال می‌شود و در نهایت با ترکیب فاز گفتار نویزی و خروجی حاصل از فیلتر، گفتار بهسازی شده



شکل ۵: منحنی‌های RMSE مرحله آموزش الف) گفتار تمیز، ب) نویز، با الگوریتم KSVD در حوزه STFT.



شکل ۶: نتایج بهسازی گفتار با معیارهای الف) PESQ، ب) SegSNR و ج) شاخص اعوجاج گفتار^۱ (SDI).

۶- نتیجه گیری

در این مقاله، روش‌های مختلف یادگیری واژه‌نامه و بازنمایی تُنک به منظور بازسازی سیگنال‌های گفتاری مورد بررسی

قرار گرفته است. در ابتدا، به طور خلاصه به مروری بر کاربردهای آموزش واژه‌نامه و بازنمایی تُنک در پردازش سیگنال، به طور خاص سیگنال گفتار پرداخته شده است.

^۱Speech Distortion Index

مورد نظر، زمان آموزش و کاربرد نهایی بازنمایی بستگی دارد. با این حال، الگوریتم K-SVD به طور کلی به عنوان روشی کارآمد و دقیق برای بازنمایی تُنک سیگنال‌های گفتاری شناخته شده است. در نهایت، به عنوان یکی از کاربردهای بازنمایی گفتار، به بررسی بهبود کیفیت گفتار در شرایط نویزی در حوزه STFT پرداخته شد. نتایج شبیه‌سازی با معیارهای PESQ، SegSNR و SDI نشان‌دهنده کارایی بالای این روش با آموزش واژه‌نامه و بازنمایی تُنک می‌باشد. به عنوان پیشنهاداتی برای تحقیقات آینده، طراحی الگوریتم‌های یادگیری واژه‌نامه ترکیبی، بررسی کاربرد بازنمایی گفتار در سایر حوزه‌ها، و توسعه روش‌های یادگیری واژه‌نامه حوزه‌های مختلف پردازش سیگنال پیشنهاد می‌شود.

سپس، الگوریتم‌های رایج یادگیری واژه‌نامه از جمله MOD و K-SVD، همچنین الگوریتم‌های جدید و کارآمد مانند RAMC و UD4-MOD که کارایی آن‌ها تاکنون با مجموعه دادگان مصنوعی آزمایش شده بودند معرفی و مورد بحث قرار گرفت. نتایج حاصل از شبیه‌سازی‌ها با معیارهای RE، MSE، SegSNR، fwSegSNR و PESQ و STOI نشان داد که استفاده از الگوریتم یادگیری واژه‌نامه K-SVD در ترکیب با الگوریتم بازنمایی تُنک در حوزه STFT، بهترین نتایج بازسازی گفتار در مقایسه با سایر الگوریتم‌ها را نشان می‌دهد. این الگوریتم نتایج مطلوبی برای بازسازی سیگنال گفتار در سایر حوزه‌ها نیز نشان داده است. انتخاب الگوریتم مناسب برای یادگیری واژه‌نامه در بازنمایی تُنک سیگنال‌های گفتاری، به عواملی مانند نوع داده، دقت

۶- مراجع

- [1] Tsaig, Yaakov, and David L. Donoho, "Extensions of Compressed Sensing," *Signal Processing* 86, no. 3 (March 2006): 549–71.
- [2] Zhao, Yongqiang, and Jingxiang Yang, "Hyperspectral Image Denoising via Sparse Representation and Low-Rank Constraint" 53, no. 1 (January 1, 2015): 296–308.
- [3] Yang, Meng, et al., "Latent Dictionary Learning for Sparse Representation Based Classification," *Computer Vision and Pattern Recognition*, June 23, 2014.
- [4] Liu, Yu, et al., "Recent Advances in Sparse Representation Based Medical Image Fusion" 24, no. 2 (April 12, 2021): 45–53.
- [5] Sharma, Pulkit, Vinayak Abrol, and Anil Kumar Sao, "Deep-Sparse-Representation-Based Features for Speech Recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25, no. 11 (November 2017): 2162–75.
- [6] Peng, Peng, et al., "Sparse Representation and Dictionary Learning Model Incorporating Group Sparsity and Incoherence to Extract Abnormal Brain Regions Associated with Schizophrenia," *IEEE Access* 8 (January 1, 2020): 104396–406.
- [7] Tournier, Jean-Yves, et al., "Sparse Representations and Dictionary Learning: From Image Fusion to Motion Estimation," *HAL (Le Centre Pour La Communication Scientifique Directe)*, July 11, 2021.
- [8] Mavaddaty, Samira, Seyed Mohammad Ahadi, and Sanaz Seyedin, "Speech Enhancement Using Sparse Dictionary Learning in Wavelet Packet Transform Domain," *Computer Speech & Language* 44 (July 2017): 22–47.
- [9] Eshaghi, Mohadesse, Farbod Razzazi, and Alireza Behrad, "A Voice Activity Detection Algorithm in Spectro-Temporal Domain Using Sparse Representation," *International Journal of Machine Learning and Cybernetics*, August 1, 2018.
- [10] Sugiura, Yosuke, and Tetsuya Shimamura, "Speech Enhancement Based on Sparse Representation in Logarithmic Frequency Scale," *2018 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, November 1, 2018.
- [11] Shaheen, Dima, Oumayma Al Dakkak, and Mohiedin Wainakh, "Incoherent Discriminative Dictionary Learning for Speech Enhancement," *Journal of Telecommunications and Information Technology* 3 (2018): 42–54.

- [12] Ji, Yunyun, Wei-Ping Zhu, and Benoit Champagne, "Speech Enhancement Based on Dictionary Learning and Low-Rank Matrix Decomposition." *IEEE Access* 7 (January 1, 2019): 4936–47.
- [13] Bai, Huang, et al., "Unit-Norm Tight Frame-Based Sparse Representation with Application to Speech Inpainting." *Digital Signal Processing* 123 (April 2022).
- [14] Wang, Lianzi, et al., "Voiced/Unvoiced Pronunciation Judgement Based on Sparse Representation and Learning Dictionary," December 1, 2019.
- [15] Ding, Shaojin, et al., "Learning Structured Sparse Representations for Voice Conversion," *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (January 1, 2020): 343–54.
- [16] Haneche, Houria, Bachir Boudraa, and Abdeldjalil Ouahabi, "A New Way to Enhance Speech Signal Based on Compressed Sensing." *Measurement* 151 (February 2020).
- [17] Mavaddati, Samira, "A New Method for Speech Enhancement Based on Incoherent Model Learning in Wavelet Transform Domain." *Pardāzish-i 'Alāim va Dādah/Hā* 17, no. 3 (November 1, 2020).
- [18] Al-Hassani, Ihsan, Oumayma Al-Dakkak, and Abdlnaser Assami, "Phonetic Segmentation Using a Wavelet-Based Speech Cepstral Features and Sparse Representation Classifier," *Journal of Telecommunications and Information Technology* 4 (December 30, 2021): 12–22.
- [19] Kwek, Lee-Chung, et al., "Sparse Representation and Reproduction of Speech Signals in Complex Fourier Basis," *International Journal of Speech Technology* 25, no. 1 (November 26, 2021): 211–17.
- [20] Kwek, Lee-Chung, et al., "Sparse Representation and Reproduction of Speech Signals in Complex Fourier Basis," *International Journal of Speech Technology* 25, no. 1 (November 26, 2021): 211–17.
- [21] Reddy, M. Kiran, and Paavo Alku, "Exemplar-Based Sparse Representations for Detection of Parkinson's Disease from Speech," *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 31 (January 1, 2023): 1386–96.
- [22] Xie, Yuan, Kan Xie, and Shengli Xie, "Underdetermined Blind Source Separation of Speech Mixtures Unifying Dictionary Learning and Sparse Representation." *International Journal of Machine Learning and Cybernetics* 12, no. 12 (August 3, 2021): 3573–83.
- [23] Chen, Xiaojun, et al., "A Novel Nonlinear Dictionary Learning Algorithm Based on Nonlinear-KSVD and Nonlinear-MOD." *Lecture Notes in Computer Science*, January 1, 2022, 167–79.
- [۲۴] نجاتی، منصور (۱۳۹۵)، بهبود کارایی مدل بازنمایی تُنک برای بازگردانی و بازسازی تصویر، رساله دکتری مهندسی برق دانشگاه صنعتی اصفهان
- [25] Cai, Shuting, et al., "A Dictionary-Learning Algorithm Based on Method of Optimal Directions and Approximate K-SVD," July 1, 2016
- [26] Deeba, Farah, and Kun She, "Lossless Digital Image Watermarking in Sparse Domain by Using K-Singular Value Decomposition Algorithm," *IET Image Processing*, January 28, 2020.
- [27] Parsa, Javad, et al., "Joint low mutual and average coherence dictionary learning," in *Proceeding of 26th European Signal Processing Conference (EUSIPCO 2018)*, 2018.
- [28] Parsa, Javad, et al., "A new algorithm for dictionary learning based on convex approximation," in *Proceedings of the 27th European Signal Processing Conference (EUSIPCO 2019)*, 2019.